Title

Misclassification of Breast Imaging Reporting and Data System (BI-RADS) Mammographic Density and Implications for Breast Density Reporting Legislation

Journal

The Breast Journal, 21(5)

ISSN

1075-122X

Authors

Gard, Charlotte C
Aiello Bowles, Erin J
Miglioretti, Diana L
et al.

Publication Date

2015-09-01

DOI

10.1111/tbj.12443

Peer reviewed

HHS Public Access

Author manuscript

Breast J. Author manuscript; available in PMC 2016 September 01.

Published in final edited form as:

Breast J. 2015 September ; 21(5): 481–489. doi:10.1111/tbj.12443.

Misclassification of Breast Imaging Reporting and Data System (BI-RADS) mammographic density and implications for breast density reporting legislation

Charlotte C. Gard, PhD[1], **Erin J. Aiello Bowles, MPH**[2], **Diana L. Miglioretti, PhD**[3,2], **Stephen H. Taplin, MD, MPH**[4], and **Carolyn M. Rutter, PhD**[5]

[1]Department of Economics, Applied Statistics, and International Business, New Mexico State University, Las Cruces, NM

[2]Group Health Research Institute, Group Health Cooperative, Seattle, WA

[3]Department of Public Health Sciences, UC Davis School of Medicine, Davis, CA

[4]National Cancer Institute, Behavioral Research Program, Division of Cancer Control and Population Sciences, Bethesda, MD

[5]RAND Corporation, Santa Monica, California

Abstract

U.S. states have begun legislating mammographic breast density reporting to women, requiring that women undergoing screening mammography who have dense breast tissue (BI-RADS density c or d) receive written notification of their breast density; however, the impact that misclassification of breast density will have on this reporting remains unclear. The aim of this study was to assess reproducibility of the four-category Breast Imaging Reporting and Data System (BI-RADS) density measure and examine its relationship with a continuous measure of percent density. We enrolled 19 radiologists, experienced in breast imaging, from a single integrated healthcare system. Radiologists interpreted 341 screening mammograms at two points in time six months apart. We assessed intra- and inter-observer agreement in radiologists' interpretations of BI-RADS density and explored whether agreement depended upon radiologist characteristics. We examined the relationship between BI-RADS density and percent density in a subset of 282 examinations. Intra-radiologist agreement was moderate to substantial, with kappa varying across radiologists from 0.50–0.81 (mean=0.69, 95% CI (0.63, 0.73)). Intra-radiologist agreement was higher for radiologists with ≥10 years experience interpreting mammograms (difference in mean kappa=0.10, 95% CI (0.01, 0.24)). Inter-radiologist agreement varied widely across radiologist pairs from slight to substantial, with kappa ranging from 0.02–0.72 (mean=0.46, 95% CI (0.36, 0.55)). Of 145 examinations interpreted as "non-dense" (BI-RADS density a or b) by the majority of radiologists, 82.8% were interpreted as "dense" (BI-RADS density c or d) by at least one radiologist. Of 187 examinations interpreted as "dense" by the majority of radiologists, 47.1% were interpreted as "non-dense" by at least one radiologist. While the examinations of

Correspondence: Charlotte C. Gard, Ph.D., Department of Economics, Applied Statistics, and International Business, College of Business, New Mexico State University, P. O. Box 30001 / MSC 3CQ, Las Cruces, NM 88003-8001, cgard@nmsu.edu, Phone: 575-646-5500, Fax: 575-646-1915.

almost half of the women in our study were interpreted clinically as having BI-RADS density c or d, only about 10% of examinations had percent density >50%. Our results suggest that breast density reporting based on a single BI-RADS density interpretation may be misleading due to high inter-radiologist variability and a lack of correspondence between BI-RADS density and percent density.

Keywords

BI-RADS density; misclassification; breast density reporting legislation; intra- and inter-radiologist agreement; percent density

Introduction

Mammographic breast density measures the amount of radiographically dense tissue in a woman's breast. The relationship between mammographic density and breast cancer risk has been investigated in over 40 studies dating back more than 30 years. A meta-analysis reported a strong linear association between percent density and breast cancer risk and found that women with breasts that are ≥75% dense have a four to six times greater risk of developing breast cancer than women with breasts that are <5% dense (1). In addition to increasing cancer risk, dense tissue can "mask" lesions on mammograms, decreasing the sensitivity of screening mammography and increasing the need for additional workup of uncertain mammographic findings (2, 3, 4).

In clinical settings, mammographic density is typically measured using the four-category American College of Radiology Breast Imaging Reporting and Data System (BI-RADS) scale (5). BI-RADS "density a" describes breasts that are "almost entirely fatty," "density b" describes breasts with "scattered areas of fibroglandular density," "density c" describes breasts that are "heterogeneously dense," and "density d" describes breasts that are "extremely dense." Percent density, which measures the percentage of the total area of a woman's breast occupied by dense tissue, is often ascertained in research settings using computer-assisted methods. Studies that categorize percent density have shown slightly stronger associations between breast density and breast cancer risk than studies that use BI-RADS density (1).

BI-RADS density was originally introduced to allow radiologists to record their level of concern that a cancer might be missed on mammography because of dense tissue (6) but has come to be used extensively in research on breast cancer risk and in research on mammography performance outcomes (2, 7, 8). Since 2009, at least seventeen states have passed breast density notification laws, and similar legislation is under consideration in many other states and in the U.S. Congress (9). The laws differ from state to state but, generally, require that women undergoing screening mammography who have dense breast tissue (BI-RADS density c or d) receive written notification of their breast density and be advised that dense tissue may be associated with an increased risk of breast cancer and may impact the accuracy of mammography. Women with dense breasts are encouraged to talk with their physicians about the possibility of additional screening. The U.S. Food and Drug Administration is expected to issue an amendment to the Mammography Quality Standards

Act that will address breast density reporting, potentially standardizing notification nationwide.

We address two concerns related to the use of BI-RADS density in clinical care and breast cancer research. First, we examine *random misclassification* due to variability in BI-RADS density measurements within and between radiologists. Most previous studies that have assessed agreement of BI-RADS density (10, 11, 12, 13) have been based on a small number of radiologists interpreting a small number of examinations. In studies involving more than two radiologists (11, 12, 13), confidence intervals around agreement estimates were either not provided or do not appear to have properly accounted for correlation among mammograms interpreted by the same reader. Second, we address the potential for *systematic misclassification*, comparing the BI-RADS density interpretations of individual radiologists to the BI-RADS density interpretations of the majority of radiologists and to a more objective semi-automated continuous measure of percent density.

Materials and Methods

We used data from a study of whether computer-assisted detection (CAD) improved radiologists' interpretative performance, conducted from 2001 to 2002. Detailed descriptions of the study design and test set are provided elsewhere (14). Briefly, we enrolled 19 radiologists, experienced in breast imaging, from six facilities in a single integrated healthcare system. Participating radiologists interpreted a test set of 341 bilateral screening mammograms during two four-hour sessions at the start of the study and, again, during two four-hour sessions roughly six months later. During each session, radiologists interpreted approximately 90 examinations without CAD (unassisted interpretations) and approximately 90 examinations with CAD (assisted interpretations). Radiologists reviewed craniocaudal and mediolateral oblique views of a woman's left and right breasts on a reviewer board, which included prior films, if available, and a summary sheet with the woman's age and family history of breast cancer. The ImageChecker M2 1000 system (version 2.2, R2 Technology) was used for assisted interpretations. The health plan's institutional review board approved this study. Radiologists provided informed consent, and a waiver of consent was granted for use of mammography examinations.

Women contributing examinations to the test set had a screening mammogram interpreted in the healthcare system between 1996 and 1998 and were enrolled in the healthcare system for at least two years following screening. The test set was constructed to include approximately twice as many examinations of women with cancer (i.e., invasive carcinoma or ductal carcinoma *in situ* within two years) as without and roughly equal numbers of non-dense (BI-RADS density a and b) and dense (BI-RADS density c and d) examinations based on the clinical interpretation, similar to clinical practice (15).

Density measures

Our data included three density measures for each examination. First, we defined clinical BI-RADS density as the maximum of the left and right BI-RADS densities as interpreted in clinical practice.

Second, in the test setting, radiologists recorded an overall BI-RADS density for each examination, for both unassisted and assisted interpretations, according to the ACR BI-RADS 3rd edition density definition (16). Radiologists received no special training regarding the BI-RADS lexicon because they had been using it since its inception and were unaware of the BI-RADS density interpretation made in clinical practice. We determined the majority report for each examination as the mode of the BI-RADS density interpretations from the test setting. Majority report was the same under the unassisted and assisted conditions for 315 of the 325 examinations for which there was one most frequent category. We, therefore, present the majority report for unassisted interpretations only.

Third, in 2008, continuous percent density was determined for the 282 examinations that were available to be digitized. Films were scanned using a Kodak Lumisys 85 scanner (Eastman Kodak, Rochester, NY) at a resolution of 87 microns/pixel for small films and 116 microns/pixel for large films. A single reader (E.A.B.) measured percent density using interactive thresholding (17, 18) with the Cumulus program (University of Toronto, Toronto, Canada). With interactive thresholding, a reader uses a computer program to view each digitized image and select a threshold brightness that distinguishes dense tissue from non-dense tissue. Percent density is calculated as the ratio of the number of pixels with brightness above the threshold level to the total number of pixels in the breast. All percent density measurements were made without knowledge of a woman's clinical density and were based on the craniocaudal view of the left breast.

Statistical analysis

We measured intra- and inter-radiologist agreement using percent agreement and kappa (19). Percent agreement is a "raw measure," which provides the percentage of interpretations for which both raters agree. Kappa is a "chance corrected measure," based on the difference between observed agreement and agreement expected due to chance; values less than 0 represent poor agreement, values 0.00–0.20 slight agreement, values 0.21–0.40 fair agreement, values 0.41–0.60 moderate agreement, values 0.61–0.80 substantial agreement, and values 0.81–1.00 almost perfect agreement (20).

We determined intra-radiologist agreement for each radiologist based on his or her unassisted and assisted interpretations and determined overall intra-radiologist agreement as the mean of the agreement measures for individual radiologists. Inter-radiologist agreement was determined separately for unassisted and assisted interpretations. For a given condition, we calculated agreement for each pair of radiologists based on examinations interpreted by both members of the pair and determined overall inter-radiologist agreement as the mean of the pairwise measures.

We calculated intra- and inter-radiologist agreement by cancer status and by radiologists' years of experience interpreting mammography, percent time devoted to breast imaging, and interpretive volume. Inter-radiologist agreement was similar for unassisted and assisted interpretations; therefore, we present results for unassisted interpretations only.

We used a bootstrap approach (21) to construct confidence intervals for estimates of agreement and differences in agreement. For the test set of 341 examinations, each bootstrap

sample was generated by randomly selecting 341 examinations *with replacement* from the original sample then randomly selecting 19 radiologists *with replacement*. The re-sampling of examinations was stratified by cancer status, to mimic the original study design (14). We calculated 95% confidence intervals based on 5,000 bootstrap samples using the bias-corrected percentile method.

We used Stata/SE 9.2 for Windows (StataCorp LP, College Station, TX) and R (version 2.8.0, R Foundation for Statistical Computing, Vienna, Austria) to plot distributions of BI-RADS density and percent density, respectively. Remaining analyses were performed using SAS 9.2 (SAS Institute, Cary, NC).

Results

Of the 19 participating radiologists, 13 (68.4%) had ≥10 years experience interpreting mammography, 12 (63.2%) devoted 20–39% of their time to breast imaging, and 10 (52.6%) interpreted ≤50 mammography examinations in an average week (Table 1). Of 341 women contributing examinations to the test set, 227 (66.6%) had a diagnosis of invasive carcinoma or ductal carcinoma *in situ* within two years following their clinical examination.

Our analyses included 341 BI-RADS density interpretations obtained in clinical practice, 6,263 interpretations under the unassisted test condition, and 6,233 interpretations under the assisted test condition. Radiologists interpreted BI-RADS density for 251–341 examinations under the unassisted test condition, 251–341 examinations under the assisted test condition, and 162–341 examinations under both test conditions. Some of the examinations in the test set had been interpreted by radiologists in our study in clinical practice. Fourteen of the 19 study radiologists interpreted examinations for women in the study in clinical practice, with the number of clinical interpretations per study radiologist ranging from 4–30.

Clinically, 5.9%, 44.0%, 38.4% and 11.7% of examinations were interpreted as having BI-RADS density a, b, c, and d, respectively (Figure 1). Overall, radiologists were more likely to assign BI-RADS density b and less likely to assign BI-RADS density c in the clinical setting than in the test setting. Across individual radiologists, the percentage of unassisted examinations interpreted as BI-RADS density a, b, c, and d ranged from 0 to 17%, 3.1 to 52.8%, 30.9 to 90.2%, and 2.4 to 28.5%, respectively. Of 145 examinations (43.7%) with majority report a or b, 120 (82.8%) were interpreted as having BI-RADS density c or d by at least one radiologist in the test setting. Of 187 examinations (56.3%) with majority report c or d, 88 (47.1%) were interpreted as having BI-RADS density a or b by at least one radiologist. Almost all examinations (n = 316, 92.7%) were interpreted as having BI-RADS density c or d by at least one radiologist.

Intra- and inter-radiologist agreement

Percent intra-radiologist agreement for individual radiologists ranged from 66 to 95%, with mean 82%, 95% confidence interval (CI) (78%, 85%), and kappa values ranged from 0.50 to 0.81, with mean 0.69, 95% CI (0.63, 0.73) (Table 2). Percent inter-radiologist agreement for radiologist pairs ranged from 33 to 82%, with mean 65%, 95% CI (59%, 71%), and pairwise kappa values ranged from 0.02 to 0.72, with mean 0.46, 95% CI (0.36, 0.55). One

experienced radiologist exhibited only slight agreement with other radiologists, with pairwise kappas in the range 0.02 to 0.19. With the removal of this radiologist's data, the minimum pairwise kappa increased to 0.07, and overall inter-radiologist agreement increased to 0.50, 95% CI (0.43, 0.57). We found no difference in intra- or inter-radiologist agreement for examinations of women with cancer versus without cancer.

Radiologists with ≥10 years experience had higher intra-radiologist agreement (mean kappa=0.72, 95% CI (0.67, 0.76)) than radiologists with <10 years experience (mean kappa=0.62, 95% CI (0.48, 0.70)). Radiologists with ≥10 years experience also had higher inter-radiologist agreement (mean pairwise kappa=0.52, 95% CI (0.41, 0.61)) than radiologists with <10 years experience (mean pairwise kappa=0.36, 95% CI (0.15, 0.54)), although this difference was not statistically significant (difference=0.16, 95% CI (-0.06, 0.40)). We observed no relationship between intra-radiologist or inter-radiologist agreement and radiologists' percent time devoted to breast imaging or interpretive volume.

We observed moderate agreement between clinical BI-RADS density and majority report, with kappa equal to 0.55, 95% CI (0.48, 0.64). The clinical interpretation and majority report agreed for 71% of examinations and rarely disagreed by more than one density category (Table 3). Examinations with clinical density a were more likely to have a majority report of b than a (55% versus 45%, respectively).

Percent density

Median percent density increased with clinical BI-RADS density, but there was considerable overlap in the range of percent density across BI-RADS density categories. Examinations with clinical BI-RADS density a, b, c, and d had percent density ranging from 1 to 20% (median 4%), 1 to 65% (median 14%), 8 to 61% (median 30%), and 26 to 91% (median 52%), respectively (Figure 2).

Of 104 examinations with clinical BI-RADS density c, 34 (32.7%) had percent density <25%, and 64 (61.5%) had percent density 25-50% (Table 4). Of 35 examinations with clinical BI-RADS density d, 13 (37.1%) had percent density 25-50%.

Across radiologists, distributions of percent density were similar for examinations interpreted as having BI-RADS density a in the test setting and for examinations interpreted as having BI-RADS density b (Figure 3). We saw more within- and between-radiologist variability in percent density distributions for examinations interpreted as having BI-RADS density c and d.

Discussion

Among 19 radiologists interpreting 341 screening examinations twice, we observed substantial overall intra-radiologist agreement, consistent with prior studies (10, 12). Radiologists exhibited moderate or better agreement in their interpretation of the same examination, agreeing with themselves 66–95% of the time. In contrast, inter-radiologist agreement was only moderate and was lower than in most previous studies (10, 12, 22). Pairs of radiologists agreed only 33–82% of the time. Radiologists with more years

experience interpreting mammograms exhibited higher agreement than radiologists with fewer years experience, but we observed no relationship between agreement and other radiologist characteristics.

More than 50% of examinations with clinical BI-RADS density a were interpreted as having BI-RADS density b by the majority of radiologists in our study, and almost 40% of examinations with clinical BI-RADS density d were interpreted as having BI-RADS density c by the majority of radiologists. This misclassification has implications for risk prediction modeling that incorporates BI-RADS density, as the predicted risk for a woman with BI-RADS density a or d (i.e., a woman at lowest or highest risk for breast cancer based on breast density) may depend upon the radiologist who interprets her examination. Women who use online tools (23) to estimate their risk of developing breast cancer may find that their risk is under- or over-estimated because of misclassification of BI-RADS density. For example, Tice et al. (24) found that, among women age 64 years and younger, five-year risk of developing invasive breast cancer for women with BI-RADS density b was two to two and a half times that for women of the same age with BI-RADS density a.

Misclassification also has implications for the reporting of breast density to women who have dense breasts. More than 80% of the women who were determined to have non-dense breasts by the majority of radiologists in our study were found to have dense breasts by at least one radiologist. Such false positive reporting of dense breasts could lead numerous women to undergo additional, possibly unnecessary, screening tests or to be unnecessarily concerned about increased risk of cancer. Almost half of the women who were determined to have dense breasts by the majority of radiologists were found to have non-dense breasts by at least one radiologist, suggesting that misclassification may also lead to false negative reporting of dense breasts, possibly creating a false sense of security among these women.

Radiologists varied in their assignment of BI-RADS density relative to percent density but tended to assign higher BI-RADS densities to examinations with greater percent densities. While the examinations of almost half of the women in our study were interpreted clinically as having BI-RADS density c or d, only about 10% of examinations had percent density >50%. This, too, has implications for the reporting of breast density to women, as a woman may be considered to have dense breasts based on the categorical density measure but not a continuous density measure.

Our study has several limitations. Study radiologists practiced within a single healthcare system and the majority had more than 10 years of interpretive experience at the time of the study. Our results may, therefore, underestimate variability in a more diverse population of radiologists. In our study, BI-RADS density was interpreted based on four views of the breast; percent density was measured based on only one view. Studies have shown a strong correlation between percent density for the left and right breasts and between the craniocaudal and mediolateral oblique views when percent density is measured using interactive thresholding (25). Still, our results may underestimate disagreement between BI-RADS density and percent density. Radiologists in our study interpreted film-screen mammograms. While use of digital mammography is becoming more common in clinical settings, Harvey et al. (26) found no difference in reported BI-RADS density for film-screen

versus digital mammograms. Finally, for comparisons of BI-RADS density and percent density, we categorized percent density according to the BI-RADS 4th edition density definition (27), which defined density both in qualitative terms and in terms of the percent glandular material in the breast (<25%, 25–50%, 51–75%, and >75% for categories a, b, c, and d, respectively). References to percent glandular material did not appear in the BI-RADS 3rd edition density definition (16) and were eliminated in the 5th edition density definition (5). We might have used other cut points to categorize percent density. However, given the considerable overlap in the range of percent density across BI-RADS density categories that we observed, our study suggests that some women identified as having dense breasts based on BI-RADS density may not be identified as having dense breasts based on percent density and vice versa, regardless of the cut points used. Among its major strengths, our study included more radiologists interpreting more examinations than most previous studies, properly accounted for correlation among examinations interpreted by the same radiologist in quantifying errors in estimates of agreement, and is one of the first to investigate how agreement depends upon radiologist characteristics.

There will likely be a place for BI-RADS density in clinical care and breast cancer research for some time to come. While fully automated and objective measures of breast density, such as the three-dimensional measures provided by digital breast tomosynthesis (28), are on the horizon, they are at least several years away from widespread use. BI-RADS density measurements are limited, however, in part because of the misclassification issues we have described. Investigators using BI-RADS density for research purposes should, therefore, attempt to understand the limitations of the measure and address these limitations in their research. For example, statistical models may be improved by including information about misclassification of BI-RADS density and how this differs by radiologist (29). When only BI-RADS density measurements are available, it may be useful to include in models a covariate indicating the radiologist who interpreted BI-RADS density or covariates representing radiologist characteristics. Our study suggests that breast density reporting legislation has the potential to impact large numbers of women undergoing screening mammography, with both false positive and false negative reporting of dense breasts possible due to misclassification of BI-RADS density. As breast density reporting is integrated into clinical practice, further consideration as to how best to notify women of their BI-RADS density, in a way that informs them of the risks associated with high density, while at the same time conveying the limitations of the measure, may be in order. Women and their physicians should understand that breast density is an imperfect measure and is expected to vary, even when an examination is interpreted two different times by the same radiologist, with potentially greater differences when interpreted by two different radiologists.

Acknowledgments

This work was supported by the National Cancer Institute-funded Breast Cancer Surveillance Consortium (BCSC) co-operative agreement (U01CA86076, U01CA63731). A list of the BCSC investigators and procedures for requesting BCSC data for research purposes are provided at: http://breastscreening.cancer.gov/.

The opinions are solely those of the authors and do not imply any endorsement by the National Cancer Institute or the federal government. In addition, R2 Technology provided equipment and technical assistance for the project. The findings are those of the authors and cannot be construed to reflect the thoughts or opinions of R2 Technology.

References

1. McCormack VA, dos Santos Silva I. Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. Cancer Epidem Biomar. 2006; 15(6):1159–69.
2. Carney PA, Miglioretti DL, Yankaskas BC, et al. Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening mammography. Ann Intern Med. 2003; 138(3):168–75. [PubMed: 12558355]
3. Kerlikowske K, Grady D, Barclay J, Sickles EA, Ernster V. Effect of age, breast density, and family history on the sensitivity of first screening mammography. J Amer Med Assoc. 1996; 276(1):33–8.
4. Buist DSM, Porter PL, Lehman C, Taplin SH, White E. Factors contributing to mammography failure in women aged 40–49 years. J Natl Cancer I. 2004; 96(19):1432–40.
5. American College of Radiology. Breast Imaging Reporting and Data System. 5. Reston, VA: American College of Radiology; 2013.
6. Yaffe MJ. Mammographic density - measurement of mammographic density. Breast Cancer Res. 2008; 10(3):209. [PubMed: 18598375]
7. Smith-Bindman R, Chu P, Miglioretti DL, et al. Physician predictors of mammographic accuracy. J Natl Cancer I. 2005; 97:358–67.
8. Elmore JG, Jackson SL, Abraham L, et al. Variability in interpretive performance at screening mammography and radiologists' characteristics associated with accuracy. Radiology. 2009; 253:641–51. [PubMed: 19864507]
9. Are you Dense Advocacy. D.E.N.S.E.® State Efforts. [accessed June 12, 2014] Available at: http://areyoudenseadvocacy.org/dense/
10. Kerlikowske K, Grady D, Barclay J, et al. Variability and accuracy in mammographic interpretation using the American College of Radiology Breast Imaging Reporting and Data System. J Natl Cancer I. 1998; 90(23):1801–09.
11. Berg WA, Campassi C, Langenberg P, Sexton MJ. Breast Imaging Reporting and Data System: Inter- and intraobserver variability in feature analysis and final assessment. Am J Roentgenol. 2000; 174(6):1769–77. [PubMed: 10845521]
12. Ciatto S, Houssami N, Apruzzese A, et al. Categorizing breast mammographic density: intra- and interobserver reproducibility of BI-RADS density categories. Breast. 2005; 14(4):269–75. [PubMed: 16085233]
13. Ooms EA, Zonderland HM, Eijkemans MJC, et al. Mammography: Interobserver variability in breast density assessment. Breast. 2007; 16(6):568–76. [PubMed: 18035541]
14. Taplin SH, Rutter CM, Lehman CD. Testing the effect of computer-assisted detection on interpretive performance in screening mammography. Am J Roentgenol. 2006; 187(6):1475–82. [PubMed: 17114540]
15. Kerlikowske K, Ichikawa L, Miglioretti DL, et al. Longitudinal measurement of clinical mammographic breast density to improve estimation of breast cancer risk. J Natl Cancer Inst. 2007; 99:386–95. [PubMed: 17341730]
16. American College of Radiology. Breast Imaging Reporting and Data System. 3. Reston, VA: American College of Radiology; 1998.
17. Byng JW, Boyd NF, Fishell E, Jong RA, Yaffe MJ. The quantitative analysis of mammographic densities. Phys Med Biol. 1994; 39(10):1629–38. [PubMed: 15551535]
18. Byng JW, Yaffe MJ, Jong RA, et al. Analysis of mammographic density and breast cancer risk from digitized mammograms. Radiographics. 1998; 18(6):1587–98. [PubMed: 9821201]
19. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas. 1960; 20:37–46.
20. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977; 33(1):159–74. [PubMed: 843571]
21. Efron, B.; Tibshirani, RJ. An Introduction to the Bootstrap. 2. New York: Chapman and Hall/CRC; 1998.
22. Gweon HM, Youk JH, Kim JA, Son EJ. Radiologist assessment of breast density by BI-RADS categories versus fully automated volumetric assessment. Am J Roentgenol. 2013; 201(3):692–97. [PubMed: 23971465]

23. Breast Cancer Surveillance Consortium. [accessed June 12, 2014] Breast Cancer Risk Calculator. Available at: https://tools.bcsc-scc.org/BC5yearRisk/intro.htm

24. Tice JA, Cummings SR, Smith-Bindman R, Ichikawa L, Barlow WE, Kerlikowske K. Using clinical factors and mammographic breast density to estimate breast cancer risk: development and validation of a new predictive model. Ann Intern Med. 2008; 148(5):337–47. [PubMed: 18316752]

25. Yaffe MJ, Boyd NF, Byng JW, et al. Breast cancer risk and measured mammographic density. Eur J Cancer Prev. 1998; 7 (Suppl 1):S47–55. [PubMed: 10866036]

26. Harvey JA, Gard CC, Miglioretti DL, et al. Reported mammographic density: film-screen versus digital acquisition. Radiology. 2013; 266(3):752–58. [PubMed: 23249570]

27. American College of Radiology. Breast Imaging Reporting and Data System. 4. Reston, VA: American College of Radiology; 2003.

28. Kontos D, Bakic PR, Carton AK, Troxel AB, Conant EF, Maidment AD. Parenchymal texture analysis in digital breast tomosynthesis for breast cancer risk estimation: a preliminary study. Acad Radiol. 2009; 16(3):283–98. [PubMed: 19201357]

29. Gustafson, P. Measurement Error and Misclassification in Statistics and Epidemiology: Impacts and Bayesian Adjustments. New York: Chapman and Hall/CRC; 2004.

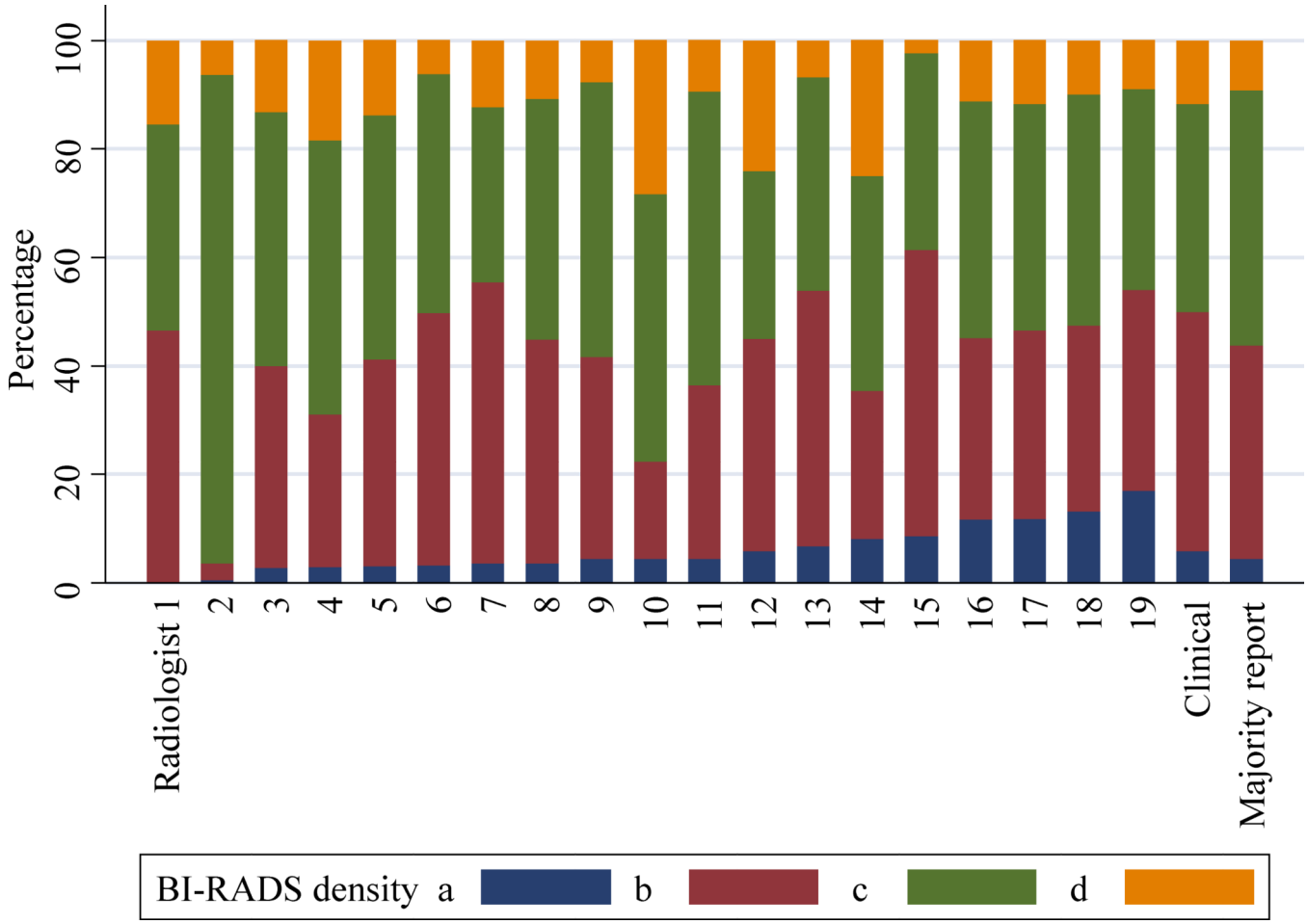

Figure 1.

Distributions of BI-RADS density interpretations from the test setting* by radiologist, clinical BI-RADS density interpretations, and majority report from the test setting*ǂ.

*Unassisted interpretations

ǂMajority report was determined for each examination as the mode of BI-RADS density interpretations from the test setting for that examination.

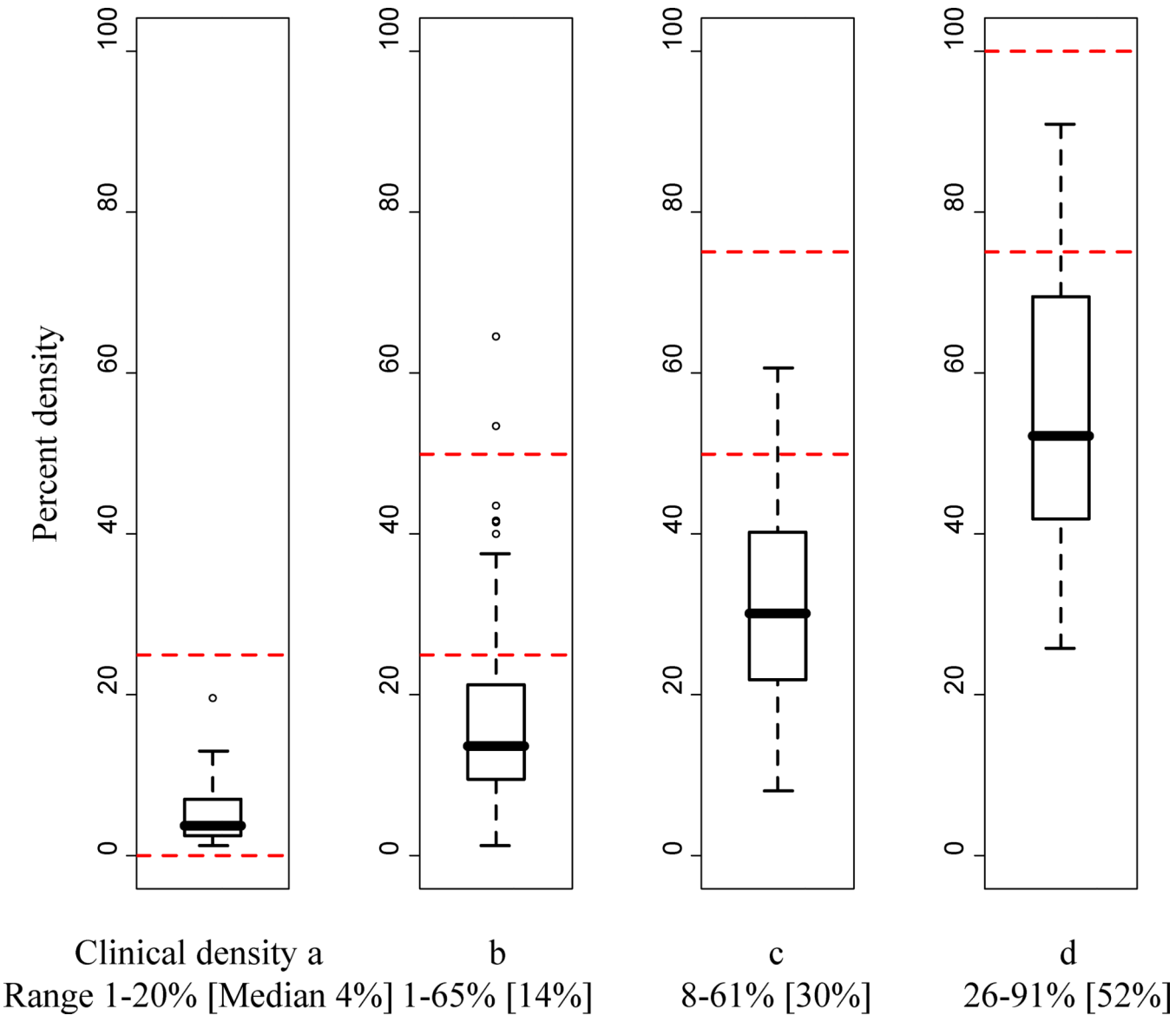

Figure 2.

Distributions of percent density within each category of clinical BI-RADS density.

The dashed horizontal lines provide the ranges of percent density that would be expected within each BI-RADS density category based on the BI-RADS 4th edition density definition. The thick horizontal line provides the median and the box provides the interquartile range for each category of clinical BI-RADS density. Whiskers extend 1.5 times the interquartile range. Outliers are also plotted.

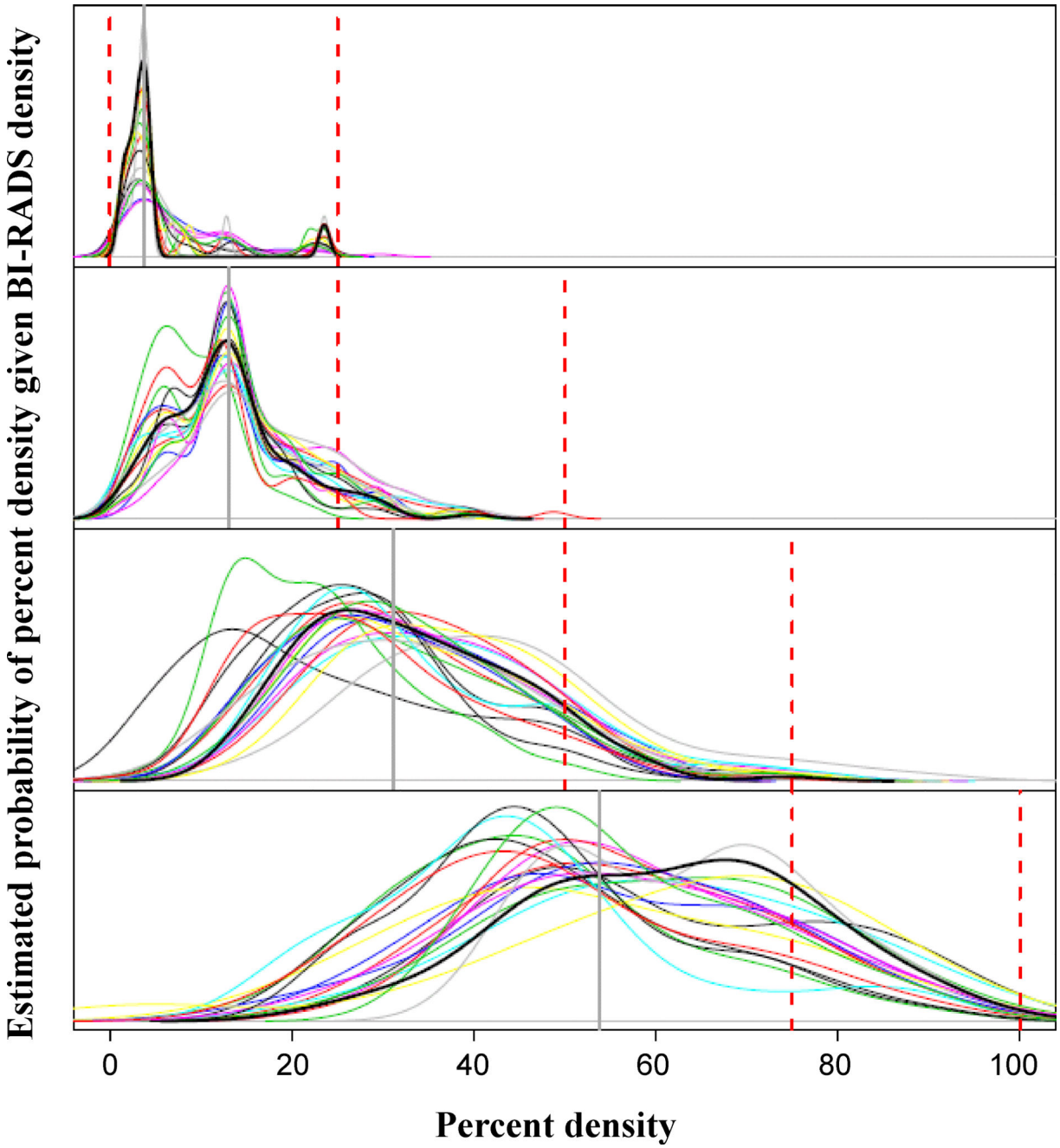

Figure 3.

Distributions of percent density for (top to bottom) BI-RADS density category a, b, c, and d interpretations* from the test setting, by radiologist.

*Unassisted interpretations

The dashed vertical lines provide the ranges of percent density that would be expected within each BI-RADS density category based on the BI-RADS 4th edition density definition. The solid vertical lines provide, for each BI-RADS density category, the overall median of percent density measurements for that category. For a given BI-RADS density category, the

overall median was calculated as the median of the median of percent density measurements for each radiologist.

Curves based on majority report are overlaid in thick black.

Table 1

Characteristics of radiologists participating in the study

Characteristic	Number of radiologists (%)
Total	19
Years experience interpreting mammography	
1–4	2 (10.5)
5–9	4 (21.1)
≥10	13 (68.4)
Percent time devoted to breast imaging	
<20%	5 (26.3)
20–39%	12 (63.2)
≥40%	2 (10.5)
Number of examinations interpreted in average week	
≤50	10 (52.6)
51–100	7 (36.8)
>100	2 (10.5)

Table 2

Intra- and inter-radiologist agreement and 95% confidence intervals (CI) overall, by cancer status, and by radiologist characteristics

	Intra-radiologist agreement				Inter-radiologist agreement*			
	Percent agreement		Kappa		Pairwise percent agreement		Pairwise kappa	
	Range	Mean (95% CI)	Range	Mean (95% CI)	Range	Mean (95% CI)	Range	Mean (95% CI)
Overall	66–95	82 (78, 85)	0.50–0.81	0.69 (0.63, 0.73)	33–82	65 (59, 71)	0.02–0.72	0.46 (0.36, 0.55)
Cancer status								
With cancer	65–94	82 (78, 85)	0.43–0.82	0.68 (0.62, 0.74)	31–85	65 (59, 71)	0.02–0.75	0.46 (0.35, 0.56)
Without cancer	68–97	81 (78, 85)	0.54–0.80	0.7 (0.63, 0.76)	33–84	65 (58, 71)	0.02–0.74	0.46 (0.36, 0.57)
Difference		0 (–3, 4)		−0.02 (−0.10, 0.06)		1 (−4, 6)		0 (−0.08, 0.07)
Years experience interpreting mammography								
<10	66–95	80 (72, 88)	0.50–0.71	**0.62 (0.48, 0.70)**	33–73	59 (44, 68)	0.06–0.57	0.36 (0.15, 0.54)
≥10	74–89	82 (79, 85)	0.62–0.81	**0.72 (0.67, 0.76)**	33–82	68 (60, 75)	0.07–0.72	0.52 (0.41, 0.61)
Difference		−3 (−11, 6)		**−0.1 (−0.24, −0.01)**		−10 (−26, 3)		−0.16 (−0.40, 0.06)
Percent time devoted to breast imaging								
<20%	66–89	80 (71, 87)	0.51–0.81	0.68 (0.56, 0.79)	54–78	68 (57, 78)	0.31–0.63	0.5 (0.35, 0.64)
≥20%	74–95	82 (79, 85)	0.50–0.78	0.69 (0.62, 0.74)	37–82	64 (57, 71)	0.08–0.72	0.44 (0.32, 0.55)
Difference		−2 (−12, 6)		−0.01 (−0.14, 0.11)		4 (−9, 15)		0.05 (−0.13, 0.23)
Number of examinations interpreted in average week								
≤50	66–87	81 (76, 85)	0.51–0.78	0.69 (0.63, 0.75)	48–81	69 (62, 76)	0.25–0.70	0.52 (0.42, 0.62)
>50	74–95	82 (78, 87)	0.50–0.81	0.68 (0.58, 0.74)	37–81	62 (53, 72)	0.08–0.70	0.41 (0.25, 0.58)
Difference		−2 (−8, 4)		0.01 (−0.08, 0.12)		7 (−6, 18)		0.11 (−0.09, 0.29)

*Unassisted interpretations

Table 3

Clinical BI-RADS density versus majority report from the test setting[*ǂ]

	BI-RADS density majority report from the test setting				
Clinical	**a**	**b**	**c**	**d**	**Total**
a	9 (45.0%)	11 (55.0%)			20 (6.1%)
b	6 (4.1%)	102 (69.4%)	38 (25.9%)	1 (0.7%)	147 (44.3%)
c		17 (13.4%)	103 (81.1%)	7 (5.5%)	127 (38.3%)
d			15 (39.5%)	23 (60.5%)	38 (11.4%)
Total	15 (4.5%)	130 (39.2%)	156 (47.0%)	31 (9.3%)	332

[*] Unassisted interpretations

[ǂ] Majority report was determined for each examination as the mode of BI-RADS density interpretations from the test setting for that examination.

Table 4

Clinical BI-RADS density versus categorized percent density in 282 examinations available to be digitized*

	Percent density				
Clinical	**<25%**	**25–50%**	**51–75%**	**>75%**	**Total**
a	18 (100.0%)				18 (6.4%)
b	104 (83.2%)	19 (15.2%)	2 (1.6%)		125 (44.3%)
c	34 (32.7%)	64 (61.5%)	6 (5.8%)		104 (36.9%)
d		13 (37.1%)	18 (51.4%)	4 (11.4%)	35 (12.4%)

*Percent density measurements are categorized according to the BI-RADS 4th edition density definition.